

Pietsch, Marcus; Ehmke, Timo

## **Die Qualität des Unterrichts an SINUS-Grundschulen. Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg**

*Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven. Münster [u.a.] : Waxmann 2015, S. 295-316. - (HANSE - Hamburger Schriften zur Qualität im Bildungswesen; 15)*



### Quellenangabe/ Reference:

Pietsch, Marcus; Ehmke, Timo: Die Qualität des Unterrichts an SINUS-Grundschulen. Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg - In: Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven. Münster [u.a.] : Waxmann 2015, S. 295-316 - URN: urn:nbn:de:0111-pedocs-115462 - DOI: 10.25656/01:11546

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-115462>

<https://doi.org/10.25656/01:11546>

in Kooperation mit / in cooperation with:



**WAXMANN**  
[www.waxmann.com](http://www.waxmann.com)

<http://www.waxmann.com>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Marcus Pietsch, Barbara Scholand,  
Klaudia Schulte (Hrsg.)

# Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:  
Grundlagen, Befunde,  
Perspektiven



HANSE  
Hamburger Schriften zur Qualität  
im Bildungswesen

herausgegeben von  
Detlef Fickermann und Knut Schwippert

Band 15



Waxmann 2015  
Münster • New York

Marcus Pietsch, Barbara Scholand,  
Klaudia Schulte (Hrsg.)

# Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:  
Grundlagen, Befunde und Perspektiven



Das gedruckte Buch finden Sie [hier](#).



Waxmann 2015  
Münster • New York

**Bibliografische Informationen der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation  
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten  
sind im Internet über <http://dnb.d-nb.de> abrufbar.

**HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 15**

ISSN 1864-2225

ISBN 978-3-8309-3278-9

© Waxmann Verlag GmbH, 2015  
Steinfurter Straße 555, 48159 Münster

[www.waxmann.com](http://www.waxmann.com)  
[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Pleßmann Design, Ascheberg  
Umschlagfoto: © Robert Kneschke – Fotolia.de  
Lektorat und Satz: Judith Zimmer, Hamburg  
Druck: Mediaprint, Paderborn

Gedruckt auf alterungsbeständigem Papier,  
säurefrei gemäß ISO 9706

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.  
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages  
in irgendeiner Form reproduziert oder unter Verwendung elektronischer  
Systeme verarbeitet, vervielfältigt oder verbreitet werden.

# Inhalt

Vorwort. ....	9
---------------	---

## Teil 1: Die Grundlagen der Schulinspektion

*Norbert Maritzen*

Schulinspektion – Aspekte der Transformation von Governance-Strukturen . . .	13
--	----

*Norbert Maritzen & Herbert Altrichter*

Schulinspektion in Hamburg – ein Gespräch über Startbedingungen und Entwicklungen . . . . .	37
--	----

*Martina Diedrich*

Aufbau und Rolle der Schulinspektion Hamburg . . . . .	57
--	----

*Thomas Brüsemeister*

Erfahrungsbereiche der Hamburger Schulinspektion – Governance-analytische Programmtheorie . . . . .	77
--	----

*Susanne Ulrich*

Das Qualitätsleitbild der Schulinspektion Hamburg . . . . .	101
---	-----

## Teil 2: Die Arbeit der Schulinspektion

*Marcus Pietsch, Ann-Katrin van den Ham & Olaf Köller*

Wirkungen von Schulinspektion: Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten . . . . .	117
---	-----

*Moritz G. Sowada*

Expertenurteile – Achillesferse oder Trumpf der Schulinspektion? . . . . .	137
--	-----

*Knut Schwippert*

Daten für die Schulentwicklung – auf die Perspektive kommt es an . . . . .	157
--	-----

<i>Barbara Scholand unter Mitarbeit von Stephanie Graw</i>	
Wie Schulinspektor/innen fragen und Grundschüler/innen antworten . . . . .	177

<i>Peter Schulze &amp; Christiane von Schachtmeyer</i>	
Der Blick aus der Praxis – die Schulformexpertinnen und -experten . . . . .	219

### **Teil 3: Die Befunde der Schulinspektion**

<i>Melanie Ehren &amp; Jaap Scheerens</i>	
Evidenzbasierte Referenzrahmen zur Schulqualität als Grundlage von Schulinspektion . . . . .	233

<i>Holger Gärtner</i>	
Zusammenhang von Schul- und Unterrichtsqualität und schulischen Rahmenbedingungen . . . . .	273

<i>Marcus Pietsch &amp; Timo Ehmke</i>	
Die Qualität des Unterrichts an SINUS-Grundschulen. Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg . . . . .	295

<i>Klaudia Schulte &amp; Markus Lücken</i>	
Der Einfluss schulischer Prozesse auf die Lernentwicklung der Schülerinnen und Schüler an weiterführenden Schulen in Hamburg . . . . .	317

<i>Susanne Peters</i>	
Eltern als Stakeholder von Schule. Erkenntnisse über die Sichtweise von Eltern durch die Hamburger Schulinspektion. . . . .	341

### **Teil 4: Die Zukunft der Schulinspektion**

<i>Sabine Müller</i>	
Zukunft der externen Evaluation – Trends und Perspektiven in Deutschland. .	369

<i>Hans-Jürgen Kuhn</i>	
Rechtliche, strukturelle und politische Rahmenbedingungen für externe Evaluation in Deutschland. . . . .	385

*Martina Diedrich*

Der zweite Zyklus der Schulinspektion Hamburg: ein Ausblick. . . . . 419

*Martin Spiewak, Angela Hellberg, Claudia Lemke, Gerrit Petrich,  
Norbert Rosenboom, Mina Zulal*

Die Perspektive der Stakeholder: Ein Interview von Martin Spiewak . . . . . 437

Autorinnen und Autoren . . . . . 455





## Vorwort

Vor rund zehn Jahren wurden die ersten Schulinspektorate in Deutschland eingeführt. Als Teil einer veränderten Steuerungsstrategie, die die Qualitätssicherung und -entwicklung von Schulen ins Zentrum setzt, sollen sie die Qualität schulischer Prozesse evaluieren, prozessuale Mindeststandards in Schule und Unterricht sichern und darüber hinaus Schulentwicklung stimulieren.

Die Hamburger Schulinspektion ist keine Inspektion der allerersten Stunde, wurde sie doch erst im Jahr 2007 offiziell eingerichtet. Zu diesem Zeitpunkt waren in vielen anderen Bundesländern schulische Evaluatorinnen und Evaluatoren bereits zwei bis drei Jahre im Feld und hatten die unterschiedlichsten Verfahren entwickelt und verschiedene Konzepte erprobt. Von ihren vielfältigen Erfahrungen mit diesem in Deutschland dahin unbekannten Instrument konnte die Hamburger Schulinspektion profitieren.

Gleichwohl hatte man in Hamburg bereits seit der Jahrtausendwende erste Konzepte zur externen Prozessevaluation an Einzelschulen in Hamburg diskutiert und entwickelt. Den Impuls hierfür gaben insbesondere die Ergebnisse der Längsschnittstudie LAU (Aspekte der Lernausgangslage und der Lernentwicklung), die seit Mitte der 1990er Jahre regelmäßig durchgeführt wurde. Es vergingen jedoch noch mehrere Jahre, bis die erste Hamburger Schule tatsächlich durch ein Inspektionssteam evaluiert wurde. Seitdem bewegt sich die Schulinspektion im Hamburger Schulsystem stets im Spannungsfeld von Administration, Praxis und Wissenschaft und muss dabei ihren eigenen sowie den Ansprüchen der verschiedenen schulischen und außerschulischen Stakeholder gerecht werden.

In ihrem ersten Zyklus, der im Jahr 2007 begann und bis zum Jahr 2013 andauerte, inspizierte die Schulinspektion 375 Hamburger Schulen. Während dieser sechs Jahre nahmen etwa 10.000 Lehrerinnen und Lehrer, gut 36.000 Eltern, rund 92.000 Schülerinnen und Schüler sowie etwa 2.000 Vertreterinnen und Vertreter Hamburger Unternehmen an den schriftlichen Befragungen der Inspektion teil. Darüber hinaus wurden fast 30.000 Unterrichtssequenzen durch die Inspektorinnen und Inspektoren beobachtet und etwa 2.000 Interviews mit Schulleitungen, Eltern- und Schülerräten sowie Lehrerkollegien und weiteren Mitarbeiterinnen und Mitarbeitern der inspizierten Schulen, aber auch mit Vertreterinnen und Vertretern von Ausbildungsbetrieben geführt.

Die Stärke des vorliegenden Bandes ist es, so meinen wir, dass er der besonderen Stellung der Schulinspektion im Gesamtgefüge des Hamburger Bildungssystems Rechnung trägt, indem er Akteure aus den verschiedenen Bereichen und mit unterschiedlichen Perspektiven zu Wort kommen lässt. Ein Teil der Beiträge ist empirisch ausgerichtet und analysiert die Arbeit der Inspektion oder nutzt die im Rahmen der Schulinspektion Hamburg erhobenen Daten für weiterführende

Analysen zum Hamburger Schulsystem. Andere Beiträge wiederum setzen sich mit den Rahmenbedingungen, Grundlagen und Folgen von Schulinspektion auseinander oder widmen sich den Erfahrungen und Erwartungen einzelner Akteure und Akteursgruppen.

Wir bedanken uns herzlich bei allen Autorinnen und Autoren dieses Bandes für die kooperative, kollegiale und anregende Zusammenarbeit. Insbesondere haben wir uns darüber gefreut, dass die vielen Kolleginnen und Kollegen aus diversen Kontexten und Arbeitsfeldern umstandslos bereit waren, ihre Expertise in Form fundierter und kritischer Beiträge sowie Gutachten beizusteuern. Sehr glücklich sind wir vor allem auch darüber, dass sich die Vertreterinnen und Vertreter der schulischen Stakeholder aktiv in diese Publikation eingebracht haben.

In der Zusammenarbeit mit all diesen verschiedenen Personen ist uns deutlich geworden, dass es institutionen- und akteursübergreifend ein großes Interesse daran gibt, über gute Bildung in und für Hamburg ins Gespräch zu kommen und sie gemeinsam zu gestalten. Dafür bieten – nicht nur, aber auch – die Schulinspektion und ihre Ergebnisse konstruktive Gesprächsanlässe. In diesem Sinne wünschen wir Ihnen eine ertragreiche Lektüre!

Hamburg im Juni 2015

Marcus Pietsch, Barbara Scholand & Klaudia Schulte

# Die Qualität des Unterrichts an SINUS-Grundschulen

## Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg

*Marcus Pietsch & Timo Ehmke*

### *Zusammenfassung*

*Der folgende Beitrag zeigt auf, wie sich Daten aus Schulinspektionsverfahren gewinnbringend für die Evaluation von Projekten und Programmen im Bildungsbereich nutzen lassen. Mithilfe eines Propensity Score Matching wird untersucht, inwieweit das weitverbreitete Unterrichtsentwicklungsprojekt SINUS einen nachweisbaren Einfluss auf die Qualität des Unterrichts an Grundschulen in Hamburg hat. Die Analysen zeigen, dass sich SINUS-Schulen in der Bewertung durch die Inspektion ausschließlich im Qualitätsbereich „Unterrichten, Lernen, Erziehen“ von vergleichbaren Schulen unterscheiden. Weiterführende Analysen von Unterrichtsbeobachtungsdaten verdeutlichen, dass die Qualität des Unterrichts an SINUS-Grundschulen rund 16 Prozent höher ausfällt als an Schulen, die unter vergleichbaren Bedingungen arbeiten. Auch werden innerschulische Qualitätsunterschiede in Folge des Programms reduziert und Schwächen im Unterricht großflächig abgestellt. Abschließend wird die praktische Bedeutung dieser Befunde für die Schul- und Unterrichtsentwicklung diskutiert.*

### 1. Einführung

Das Schul- und Unterrichtsentwicklungsprojekt SINUS gehört zu den erfolgreichsten und weitverbreitetsten Schulentwicklungsprogrammen der letzten Jahrzehnte in Deutschland (vgl. Prenzel et al. 2009). Das Akronym SINUS steht dabei für die Abkürzung „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“. Das Programm wurde 1998 in einer länderübergreifenden Initiative als ein Modellversuchsprogramm der Bund-Länder-Kommission (BLK) konzipiert und mit 180 Schulen der Sekundarstufe I in 15 Ländern gestartet (vgl. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1997). Ziel des Programms ist es, den mathematisch-naturwissenschaftlichen Unterricht weiterzuentwickeln und die Arbeit von Lehrkräften weiter zu professionalisieren. Mit dem Nachfolgeprogramm „SINUS-Transfer“ fand ab 2003 eine beachtliche Dissemination des Programms statt: Rund 1.800 Schulen und ca. 6.000 Lehrkräfte

waren daran beteiligt (vgl. Fischer et al. 2013). Ab 2004 wurde der SINUS-Ansatz auch auf die Grundschule übertragen (vgl. Fischer et. al 2009). In zehn Modulen (siehe Abschnitt 2.1) arbeiteten Lehrkräfte von ca. 400 Grundschulen aus zehn Bundesländern daran, den Mathematikunterricht und den naturwissenschaftlichen Sachunterricht in der Primarstufe weiterzuentwickeln. In einer Erweiterungsphase ab 2009 wurden die Module noch einmal neu unter bestimmten thematischen Gesichtspunkten zusammengefasst und das Programm unter der Bezeichnung „SINUS an Grundschulen“ weitergeführt (vgl. Fischer et al. 2013). Das SINUS-Programm in der Sekundarstufe I wurde ab August 2007 nicht mehr zentral koordiniert, sondern durch die einzelnen Länder eigenverantwortlich weiter koordiniert. Ab 2013 gilt dies auch für das Programm „SINUS an Grundschulen“.

Um die Qualität der Arbeit in dem Programm zu sichern, wurden von Anfang an eine wissenschaftliche Begleitforschung installiert und zahlreiche Einzelstudien durchgeführt (vgl. Abschnitt 2.2). Evaluationsstudien im Sinne der Schuleffektivitätsforschung, bei denen weitere Vergleichsgruppen, Schülerleistungen und -einstellungen oder Unterrichtsmerkmale als Erfolgskriterien herangezogen werden, gab es allerdings bislang nur wenige. Hier setzt diese Studie an und nutzt die Daten der Hamburger Schulinspektion, um die Effektivität des Schulentwicklungsprogramms SINUS an Hamburger Schulen im Hinblick auf die Unterrichtsqualität abzuschätzen.

In diesem Beitrag wird zunächst der konzeptionelle Ansatz des SINUS-Projektes dargestellt, bevor dann Forschungsansätze und bisherige Befunde zur Effektivität des SINUS-Programms zusammengefasst und die Fragestellungen für diese Studie abgeleitet werden. Die Ergebnisse unserer Analysen werden dann für Beobachtungsmerkmale im Bereich Unterricht dargestellt und im Hinblick auf ihre Bedeutung für Schulentwicklungsprogramme und mögliche Konsequenzen diskutiert.

## 2. Theoretischer Hintergrund

### 2.1 SINUS als Schul- und Unterrichtsentwicklungsprojekt

Das BLK-Programm SINUS zielt darauf ab, den Unterricht im Hinblick auf konkrete Problemstellungen weiterzuentwickeln und zu verbessern, etwa im Hinblick auf die Aufgabenkultur oder das naturwissenschaftliche Experimentieren im Unterricht. Eine besondere Rolle spielt dabei die Kooperation zwischen den Lehrkräften: In Schulverbünden, den sogenannten Sets, arbeiten Lehrerinnen und Lehrer unter wissenschaftlicher Begleitung an vorgegebenen Themenbereichen (den Modulen). Reflexion und Evaluation des eigenen Unterrichts sind dabei zentrale Elemente. Die Schulsets werden von Koordinatorinnen und Koordinatoren betreut, die eng

auf Länder- und Bundesebene zusammenarbeiteten (vgl. Prenzel et al. 2009). Ausgangspunkt der SINUS-Arbeit in der Grundschule waren zehn empirisch ermittelte Module, die sich auf typische Herausforderungen des Mathematikunterrichts und naturwissenschaftlichen Sachunterrichts in der Grundschule beziehen. Es geht dabei um folgende Schwerpunkte (vgl. Demuth et al. 2011):

- „Gute Aufgaben“,
- „Entdecken, Erforschen, Erklären“,
- „Schülervorstellungen aufgreifen, grundlegende Ideen entwickeln“,
- „Lernschwierigkeiten erkennen – verständnisvolles Lernen fördern“,
- „Talente entdecken und unterstützen“,
- „Fächerübergreifend und fächerverbindend unterrichten“,
- „Interessen von Mädchen und Jungen aufgreifen und weiterentwickeln“,
- „Eigenständig lernen – Gemeinsam lernen“,
- „Lernen begleiten – Lernergebnisse beurteilen“ und
- „Übergänge gestalten“.

Diese Module werden im Fortsetzungsprojekt „SINUS an Grundschulen“ neu zusammengefasst und auf aktuelle Fragestellungen und Herausforderungen fokussiert (vgl. Fischer et al. 2011). Daraus ergeben sich fünf Arbeitsschwerpunkte:

- „Daten aus Unterrichtsevaluationen auswerten und für die Unterrichtsentwicklung nutzen“,
- „Bildungsstandards umsetzen (in Mathematik)“,
- „Wege zu den Naturwissenschaften eröffnen (im Sachunterricht)“,
- „Kinder mit besonderen Schwierigkeiten oder besonderen Begabungen unterstützen“ und
- „Übergänge gestalten“.

## 2.2 Forschungsansätze und Befunde zur Effektivität des SINUS-Programms

Bei Unterrichtsentwicklungsprogrammen wie SINUS sollen sich die Wirkungen auf der Ebene von Schulen und Lehrpersonen in einem veränderten Unterricht und in einer positiveren Kompetenzentwicklung der Lernenden zeigen (vgl. Fischer et al. 2013). Internationale Forschungen zeigen, dass wissenschaftsbasierte Schulentwicklungsmaßnahmen diesbezüglich kurzfristig kleine, aber dennoch beobachtbare Effekte nach sich ziehen, die sich im Zeitverlauf deutlich stabilisieren und mit zunehmender Programmdauer weiter anwachsen (vgl. Bormann et al. 2003). Deutschsprachige Forschungsergebnisse hierzu liegen bislang nur vereinzelt vor

und zu dem Programm SINUS existieren bisher erst zwei Studien, die die Ergebnisse von SINUS-Schulen hinsichtlich der fachlichen Leistung von Schülerinnen und Schülern und mit Blick auf Merkmale des Unterrichts mit anderen Schulen vergleichen, die nicht am SINUS-Programm teilgenommen haben.

In der ersten Studie wurden im Rahmen einer summativen Evaluation der ersten SINUS-Welle die Wirkungen des Programms untersucht, indem die Test- und Fragebogenergebnisse von Schülerinnen und Schülern an SINUS-Schulen mit denen einer repräsentativen Schulstichprobe aus PISA 2003 verglichen wurden (vgl. Prenzel et al. 2005). Die Ergebnisse des Vergleichs zwischen SINUS-Schulen und PISA-Schulen, die nicht an dem Programm teilgenommen haben, zeigten, dass SINUS im Verlauf der ersten Programmzeit auf allen untersuchten Ebenen Wirkung entfaltet hat. Dies betraf die erfolgreiche Umsetzung der Projektinhalte auf Seiten der Lehrkräfte und die positive Wahrnehmung des Unterrichts auf Seiten der Schülerschaft. Hinsichtlich des Ziels, die Kompetenzen sowie die Interessen und Haltungen der Schülerinnen und Schüler in Mathematik und den Naturwissenschaften zu fördern, zeigten sich klare Vorteile zugunsten von SINUS bei Schulen mit mehreren Bildungsgängen und Integrierten Gesamtschulen. Hier berichteten die SINUS-Schülerinnen und -Schüler über ein höheres Interesse und mehr Freude an Mathematik.

Die Studie von Dalehefte et al. (2014) untersuchte die Wirksamkeit des Programms SINUS an Grundschulen und analysierte, inwieweit sich die mathematischen Kompetenzen von Schülerinnen und Schülern sowie das Weiterbildungs- und Kooperationsverhalten von Mathematiklehrkräften an SINUS-Grundschulen von den Ergebnissen der für Deutschland repräsentativen „Trends in International Mathematics and Science Study“ (TIMSS) 2011 unterscheiden. Der Vergleich ergab, dass Mathematiklehrkräfte an SINUS-Grundschulen in allen betrachteten Bereichen häufiger an Fortbildungen teilnehmen als Lehrkräfte, die nicht an SINUS-Schulen unterrichten. Bei Vergleich der mathematischen Kompetenzen von Schülerinnen und Schülern an SINUS-Schulen vs. anderen Schulen, ergab sich eine Effektstärke von  $d = .18$  zugunsten des SINUS-Modellprogramms. Diese entspricht in dieser Altersstufe etwa dem Lernzuwachs von einem Drittel eines Schuljahres (vgl. Hill et al. 2008). Ein Vergleich der Perzentilbänder der mathematischen Kompetenz zwischen den beiden Gruppen ergab, dass von SINUS an Grundschulen insbesondere die schwächeren Schülerinnen und Schüler profitieren.

Die bisherige SINUS-Begleitforschung hat darüber hinaus unter anderem die Rolle der Schulleitung (vgl. Fischer 2005) und Aspekte der Kooperation von Lehrkräften (vgl. Fischer & Prenzel 2007; Ostermeier 2005) und ihre Arbeit mit dem Fachgruppen-Portfolio und anderen Dokumentationsmethoden untersucht (vgl. Stadler & Meentzen 2010; Meentzen 2009; Trepke et al. 2013). Die bislang vorliegenden Befunde deuten dabei darauf hin, dass sich SINUS-Schulen von anderen Schulen im Bezug auf effektivitätsgenerierende Merkmale vor allem mit Blick auf

den Unterricht unterscheiden könnten. So berichten Dalehefte et al. (2014), dass sich SINUS-Lehrkräfte zwar hinsichtlich unterrichtsrelevanter Fortbildungen, an denen sie teilnehmen, deutlich von anderen Lehrkräften unterscheiden, sich ihr Kooperationsverhalten vor Ort jedoch nicht von dem in Kollegien unterscheidet, deren Schulen nicht am Programm teilnehmen. Gleichwohl konnte in dieser Studie nur das Fortbildungsverhalten untersucht werden. Ob dieses jedoch mit einer veränderten Unterrichtspraxis einhergeht, ist derzeit unklar.

Zusammenfassend kann festgehalten werden, dass die genannten Studien durchaus Evidenz für die Wirksamkeit des SINUS-Programms belegen. Eine Studie, die jedoch gezielt den Nachweis von kausalen Effekten von SINUS-Schulen im Vergleich zu Nicht-SINUS-Schulen durchführt und damit mit einem Vergleichsgruppen-Design im Sinne der Kausalitätsforschung (vgl. Morgan & Winship 2007) arbeitet, steht jedoch noch aus.

### 3. Forschungsfragen und Hypothesen

Mit der aktuellen Studie soll die Wirksamkeit von „SINUS an Grundschulen“ in Hamburg untersucht werden, indem mittels der Methode des *Propensity Score Matchings* eine Vergleichsgruppe von Schulen gebildet wird, die nicht an dem Schulentwicklungsprogramm teilgenommen hat und als post-hoc generierte Kontrollgruppe für Vergleiche herangezogen werden kann. Konkret wird folgenden zwei Fragestellungen nachgegangen:

- Inwieweit unterscheiden sich SINUS-Grundschulen und Grundschulen, die nicht am SINUS-Programm teilgenommen haben, hinsichtlich unterrichtlicher, erzieherischer und organisatorischer Prozessmerkmale?
- Lassen sich zwischen SINUS-Grundschulen und Grundschulen, die nicht am SINUS-Programm teilgenommen haben, Unterschiede in der Unterrichtsqualität feststellen?

Da das Schulentwicklungsprogramm SINUS an Grundschulen explizit auf die Verbesserung des mathematisch-naturwissenschaftlichen Unterrichts abzielt, sind hier tendenziell positive Effekte zu erwarten. Die bisherigen Ergebnisse der SINUS-Begleitforschung belegen – wie gezeigt – positive Effekte auf Schülerleistungen, die den Rückschluss auf eine effektivere Unterrichtspraxis an SINUS-Schulen nahelegen. Allerdings weiß man aus der Schuleffektivitätsforschung, dass hinsichtlich der Veränderung von Unterrichtsmerkmalen durch Schulentwicklungsprogramme eher kleine Effektstärken zu erwarten sind (vgl. Seidel & Shavelson 2007).



## 4. Methodisches Vorgehen

Um die Wirksamkeit des Projektes SINUS auf schulische Prozessmerkmale an Hamburger Schulen zu evaluieren, werden im Folgenden Schulen, die an dem Projekt teilnahmen, mit solchen verglichen, die unter ähnlichen Bedingungen arbeiten, jedoch nicht an SINUS teilgenommen haben. Zum Einsatz kommt hier ein *Propensity Score Matching*. Das heißt, es werden kausale Effekte nach der Kontrolle von Variablen, die sowohl die abhängige Variable als auch den Treatmentstatus beeinflussen, geschätzt. Hierfür werden in einem ersten Schritt Paare von beobachteten Einheiten gebildet, die sich hinsichtlich der kontrollierten Variablen möglichst ähnlich sind, aber unterschiedlichen Treatmentgruppen angehören. Diese gepaarten Einheiten werden dann im zweiten Schritt zur Schätzung der Wirkung des Treatments verwendet. So werden nur Einheiten miteinander verglichen, die sich zwar im Treatmentstatus unterscheiden, aber in Hinblick auf potenziell beeinflussende Variablen möglichst ähnlich sind. Als abhängige Variable wird die Qualität des Unterrichts genutzt, die im Rahmen der Schulinspektion Hamburg mittels 20-minütiger Beobachtungen erfasst wird.

### 4.1 Datengrundlage

#### *4.1.1 Messung von Unterrichtsqualität im Rahmen der Schulinspektion Hamburg*

Die Daten zur Messung von Unterrichtsqualität wurden mittels Beobachtung von Unterrichtssequenzen erhoben. Für diese Unterrichtsbeobachtungen standen in der Regel zwei bis drei Tagen pro Schule und je Sequenz 20 Minuten Beobachtungszeit zur Verfügung. Die Auswahl der zu besuchenden Unterrichtssequenzen erfolgte jeweils vor dem Schulbesuch in Form einer stratifizierten Zufallsstichprobe, wobei Beobachtungssequenzen je Schulstunde gezogen wurden. Um dieses Vorgehen zu ermöglichen, reichten die Schulen kurz vor dem Schulbesuch ihre Stundenpläne bei der Schulinspektion Hamburg ein. Aus den vorliegenden Zeiteinheiten – von der ersten bis letzten Schulstunde – wurden dann für je zwei Unterrichtsstunden dreimal so viele Sequenzen zufällig gezogen, wie Inspektorinnen und Inspektoren am Schulbesuch beteiligt waren (in der Regel drei), sodass jeder Schulinspektor resp. jede Schulinspektorin innerhalb von zwei Schulstunden drei Sequenzen beobachtete. Die Unterrichtsstunden wurden hierfür in einen ersten und einen zweiten Teil eingeteilt, sodass die Mitarbeiterinnen und Mitarbeiter der Schulinspektion Hamburg entweder zu Beginn oder zum Abschluss (bei Doppelstunden ggf. auch in der Mitte) einer Unterrichtsstunde Informationen zur Unterrichtsqualität erfassten. Im Rahmen der Stichprobenziehung und Berichtlegung geht die Schulinspektion

Hamburg davon aus, dass die Qualität von Unterricht nicht ausschließlich von der Lehrperson abhängig ist, sondern als Angebot-Nutzungs-Beziehung definiert werden muss, die durch Lehrkräfte und Schülerschaft koproduziert wird. Dies bedeutet für die Stichprobenziehung praktisch, dass Lehrkräfte im Rahmen einer Schulinspektion nicht als stichprobenrelevantes Stratum betrachtet werden, da einzelne Lehr-Lern-Settings beobachtet werden, die eine jeweils spezifische Interaktion aus Lehrkraft, Schülerschaft, Fach, Lage der Unterrichtssequenz im Tagesablauf und weiteren Merkmalen darstellen. Entsprechend können einzelne Lehrkräfte auch häufiger, jedoch in verschiedenen Kontexten und von verschiedenen Inspektionsmitgliedern gesehen werden. Die Zuweisung der Beobachter zu den zu beobachtenden Unterrichtssequenzen erfolgte ebenso zufällig wie die Ziehung der Unterrichtssequenzen. Rund 10 Prozent aller Sequenzen, die je Schule beobachtet wurden, wurden darüber hinaus aus Gründen der Qualitätssicherung durch zwei Inspektionsmitglieder gleichzeitig eingesehen. An zu inspizierenden Grundschulen werden in der Regel mindestens 40 Sequenzen betrachtet, um zu gewährleisten, dass die Aussagen zur Qualität des Unterrichts einer Schulwoche auf Ebene der Einzelschule mit 95-prozentiger Wahrscheinlichkeit bei einer ungefähr 90-prozentigen Präzision ermöglicht werden.

Zur Bestimmung der Unterrichtsqualität werden die einzelnen Sequenzen von den Inspektorinnen und Inspektoren jeweils anhand eines Bewertungsbogens beurteilt, der 30 Kriterien umfasst. Die Kriterien dienen als Indikatoren für Qualitätsmerkmale des Hamburger Orientierungsrahmen Schulqualität (Behörde für Bildung und Sport 2006) und orientieren sich primär an den Merkmalen guten Unterrichts nach Helmke (2006). Die eingesetzten Items sollen die Messung von Unterrichtsgelingensbedingungen auf Basis einer Angebot-Nutzungs-Beziehung ermöglichen. Dabei decken die eingesetzten Items, wie Stralla (2009) im Rahmen einer vergleichenden Untersuchung aufzeigt, die national und international gängigen Kriterienkataloge zur Qualität von Unterricht differenziert ab. Die 30 Items wurden auf einer vierstufigen Ratingskala (Skalenniveau: ‚trifft nicht zu‘ bis ‚trifft zu‘) bewertet, wobei eine fünfte Kategorie markiert werden konnte, sofern die Unterrichtsbeobachter ein Merkmal für ‚nicht beobachtbar‘ hielten.

#### *4.1.2 Umgang mit Daten zur Unterrichtsqualität der Schulinspektion Hamburg*

Die Schulinspektion Hamburg meldet sowohl die Summenscores einzelner Items als auch den Wert einer Gesamtskala „Unterrichtsqualität“ an die inspizierten Schulen zurück (vgl. Pietsch et al. 2011). Verschiedene Analysen von Daten der Hamburger Schulinspektion haben jedoch gezeigt, dass die eindimensionale, IRT-basierte Skala verschiedene inhaltliche und methodische Vorteile gegenüber der

Nutzung einzelner Items hat: Einerseits zeigte sich, dass die Gesamtskala ein deutlich robusterer Indikator für die Unterrichtsqualität auf Schulebene ist als einzelne Items, da hier a) eine größere Beobachterübereinstimmung nachgewiesen werden kann und b) der durch die Ebene der Schule erklärte Varianzanteil höher ausfällt (vgl. Pietsch et al. 2011). Andererseits spielen bei einer derartigen Aggregation der Beobachtungsdaten selbst hohe Datenausfälle, die durch die Nicht-Beobachtbarkeit von Kriterien zustande kommen können, nahezu keine Rolle, da durch die Behandlung der fehlenden Werte mithilfe von Marginal-Maximum-Likelihood-Methoden potenzielle Verzerrungen in der Schätzung von Parametern auf Populationsebene ausgeglichen werden können (vgl. Pietsch 2011). Dabei liegt die interne Konsistenz der Skala bei 0.928 und es lässt sich nahezu keine schulformspezifische, jedoch eine auf die jeweils untersuchten Schulen zurückführende Varianz nachweisen (vgl. Pietsch 2010). Ebenfalls ist bekannt, dass die Unterrichtsqualität, die auf diese Weise im Rahmen der Inspektion erhoben wird, positiv mit Leistungsdaten zusammenhängt (vgl. IfBQ 2012), ein Indikator für externe Validität.

Eine Besonderheit des ersten Inspektionszyklus der Schulinspektion Hamburg lag darüber hinaus darin, dass die Inspektion bei der Modellierung von Unterrichtsqualität an die von Meyer (vgl. Meyer & Klapper 2006; Meyer 2008) initiierte Diskussion von Unterrichtsstandards anknüpfte. Um anschauliche Rückmeldungen geben zu können, wurde die metrische Skala in Abschnitte geteilt, die sich inhaltlich beschreiben lassen (vgl. Pietsch 2010). Die Schulinspektion Hamburg (vgl. IfBQ 2012) selbst unterscheidet dabei vier Stufen voneinander und definiert, dass Unterricht, in dem es ausschließlich darum geht, das Lernklima und pädagogische Strukturen zu sichern (Stufe 1), den Mindestanforderungen an lernwirksamen Unterricht nicht genügt, und benennt darüber hinaus die weiteren drei Stufen als Mindeststandard (Stufe 2), Regelstandard (Stufe 3) und Maximalstandard (Stufe 4). Dabei korrespondieren diese einzelnen Stufen mit den Basisdimensionen effektiven Unterrichts – Klassenführung, Schülerorientierung, kognitive Aktivierung und Lernautonomie (vgl. Klieme 2011). Den Befunden der Schulinspektion zufolge genügen rund 7 bis 8 Prozent aller Unterrichtsstunden an Hamburgs Schulen regelmäßig nicht dem Mindeststandard, sind also durch eine nicht gelingende Klassenführung gekennzeichnet, wohingegen im Mittel rund 25 Prozent aller Stunden den Maximalstandard gewährleisten und somit das im Hamburger Orientierungsrahmen Schulqualität definierte Ideal von Unterricht erreichen.

## 4.2 Beschreibung des Datensatzes

SINUS soll zu einer Verbesserung der Unterrichtskultur an Schulen führen. Im Fokus des Programms steht insofern nicht nur der einzelne Unterricht einer spezifischen Lehrkraft innerhalb einer Schule, sondern vielmehr die Weiterentwicklung

des Unterrichts auf Schulebene (vgl. Meyer 2011). Entsprechend nehmen Schulen an dem Programm teil, nicht einzelne Lehrerinnen und Lehrer. Diese Tatsache ermöglicht es, anhand von Daten der Schulinspektion Hamburg post-hoc herauszufinden, ob eine Schule am Programm teilgenommen hat oder nicht. So wird im Rahmen der Schulinspektion grundsätzlich erhoben, an welchen Maßnahmen und Programmen eine Schule in den drei Jahren vor einer Inspektion teilgenommen hat. Für jedes Jahr vor der Inspektion kann eine separate Angabe gemacht werden. Mithilfe dieser Daten wurden in einem ersten Schritt solche Schulen identifiziert, die in den drei Jahren vor einer Inspektion mindestens ein Jahr am SINUS-Projekt teilgenommen haben. Insgesamt nahmen 49 Grundschulen in Hamburg im Vorfeld einer Inspektion mindestens ein Jahr lang am Projekt SINUS teil. An diesen Schulen wurden zum Zeitpunkt der Inspektion 6.429 Schülerinnen und Schüler von 1.239 Lehrkräften unterrichtet. Im Vergleich hierzu nahmen 113 Schulen nicht an SINUS teil. An diesen Schulen wurden zum Zeitpunkt der Inspektion 13.698 Schülerinnen und Schüler von 2.608 Lehrerinnen und Lehrern unterrichtet. Insgesamt wurden an allen Grundschulen zusammen 8.361 Unterrichtssequenzen durch Schulinspektorinnen und Schulinspektoren eingesehen. Hiervon stammen 6.012 Unterrichtsbeobachtungen aus Schulen, die laut Eigenauskunft nicht an SINUS teilgenommen haben, und 2.349 aus SINUS-Schulen. Im Gesamtdatensatz liegt die Inter-Rater-Reliabilität der Beobachter, gemessen als Intra-Klassen-Korrelation (ICC), bei 0,92, was für eine hohe Konsistenz der Beobachter spricht. Dabei liegt die Unterrichtsqualität der SINUS-Schulen (Mittelwert=0,56) auf der Skala der Schulinspektion um rund 25 Prozent (Cohen's  $d=0,18$ ) über derjenigen von Grundschulen ohne Teilnahme am Programm (Mittelwert=0,43), wobei der Unterschied hoch signifikant ist ( $p<0,001$ ).

### 4.3 Analyseverfahren

Augenscheinlich unterscheidet sich somit die Qualität des Unterrichts an SINUS-Schulen von demjenigen an Nicht-SINUS-Schulen positiv und es gelingt entsprechend, Unterricht lernwirksamer zu gestalten. Gleichwohl kann nicht garantiert werden, dass sich SINUS-Schulen von den anderen Schulen mit Blick auf mögliche Kovariaten nicht systematisch unterscheiden, da SINUS-Schulen dem Programm nicht zufällig zugeordnet werden, sondern freiwillig am Programm teilnehmen. Insofern kann der berichtete Unterschied in der Unterrichtsqualität ebenso gut Selektionsverzerrungen geschuldet sein wie einem tatsächlichen Effekt des Programms. Die Kausalaussage „SINUS führt zu einer Effektivierung des Unterrichts“ ist daher nicht haltbar. Dieser Problematik kann auf verschiedenen Wegen begegnet werden, wobei es stets darum geht, kontrafaktische Kausalanalysen post-hoc mithilfe statistischer Verfahren zu ermöglichen (vgl. z. B. Legewie 2012).

Im vorliegenden Fall wird hierfür das sogenannte *Propensity Score Matching* genutzt (vgl. Rosenbaum & Rubin 1983). Die Grundidee dieses Verfahrens besteht darin, für alle Schulen der Treatmentgruppe (SINUS-Schulen) eine Kontrollgruppe zu ermitteln, die diesen Schulen in relevanten Merkmalen möglichst ähnlich ist. Auf diesem Wege soll sichergestellt werden, dass die zentrale Annahme der kontrafaktischen Kausalanalyse, die Annahme der konditionalen Unabhängigkeit (*Conditional Independence Assumption*, CIA) erfüllt wird. Dies bedeutet, dass nach der Kontrolle von Kovariaten die Verteilung der Analyseeinheiten über Treatment- und Kontrollgruppe möglichst zufällig ist.

Die Schätzung der *Propensity Scores* sowie das Matching wurden mithilfe des SPSS-Plugins *PS Matching* (vgl. Thoemmes 2014), das auf das R-Paket *match-it* zurückgreift, durchgeführt. Für das praktische Vorgehen existieren hier verschiedene Möglichkeiten (vgl. Thoemmes & Kim 2011), wobei für die nachfolgenden Analysen ein *Nearest Neighbor Matching* mit einer 1-zu-1-Zuordnung gewählt wurde. Entsprechend wurde jeder Treatment-Schule diejenige Schule mit dem ähnlichsten *Propensity Score* zugeordnet. Dabei wurde keine Begrenzung der Distanz der Scores berücksichtigt, also kein Caliper durchgeführt. Darüber hinaus wurde die Prozedur ohne Zurücklegen durchgeführt. Nicht zugeordnete Schulen ohne SINUS-Teilnahme wurden hingegen aus dem Datensatz entfernt, sodass – wie beabsichtigt – ebenso viele Schulen als Kontroll- wie als Treatmentgruppe im Analysedatensatz verblieben.

Als zentrale Hintergrundvariablen für das Matching wurden drei Bereiche definiert, die für eine Verletzung der CIA verantwortlich sein können, für die also ein Zusammenhang mit der zentralen Analysevariable Unterrichtsqualität zu erwarten ist: a) die soziale Zusammensetzung der Schülerschaft an der Schule, gemessen an deren Sozialindex, b) die Schulgröße, indiziert über die Anzahl der Schülerinnen und Schüler an der Schule sowie die Größe der jeweiligen Lehrerkollegien und c) das Inspektionsjahr. Die Auswahl dieser Variablen ist sowohl theoretisch als auch empirisch begründet. So ist ein nicht-linearer Zusammenhang zwischen der Unterrichtsqualität, wie sie die Schulinspektion Hamburg misst, und der sozialen Zusammensetzung der Schülerschaft an Schulen empirisch nachgewiesen (vgl. IfBM 2011). Darüber hinaus ist bekannt, dass die Schulgröße sowohl einen Einfluss auf die Lernergebnisse von Schülerinnen und Schülern als auch auf Innovations- und Entwicklungsmöglichkeiten von Schulen hat (vgl. Leithwood & Jantzi 2009). Und zu guter Letzt besteht für Schulen in Hamburg durch die zeitlich versetzte Inspektion das Problem, dass Schulen im zeitlichen Verlauf eine zunehmend höhere Wahrscheinlichkeit haben, Teil der jährlichen Inspektionsstichprobe zu werden (vgl. Pietsch et al. 2014). Die Antizipation dieser Tatsache könnte beispielsweise Schulen, die in den ersten Jahren der Inspektion noch nicht inspiziert wurden, dazu bringen, dass sie mit zunehmender Zeitdauer in Erwartung eines Schulinspektionsbesuches präventiv Maßnahmen einleiten, die zu einer Verbesserung der Unter-

richtsqualität führen, welche jedoch grundsätzlich unabhängig davon sind, ob eine Schule an SINUS teilgenommen hat oder nicht.

## 5. Befunde

### 5.1 Befunde zur Matchingqualität

Mithilfe des beschriebenen Verfahrens konnten den 49 Hamburger SINUS-Grundschulen 49 vergleichbare Schulen zugeordnet werden. Nach dem Matching stehen den 2.349 beobachteten Unterrichtssequenzen an SINUS-Schulen 2.533 Sequenzen an vergleichbaren Schulen gegenüber. Im Mittel wurden an Treatment-Schulen rund 48 Sequenzen von der Inspektion eingesehen, an Schulen der Kontrollgruppe 52. Die Güte des Matchings lässt sich dabei anhand verschiedener Indikatoren überprüfen: Einerseits sollten sich die beiden Gruppen hinsichtlich der kontrollierten Variablen nach dem Matching möglichst nicht mehr statistisch nachweisbar unterscheiden. Andererseits sollte der Standardisierte Bias, der Unterschied in den Mittelwerten der Kovariaten unter Berücksichtigung der Varianz dieser Merkmale (vgl. Rosenbaum & Rubin 1985), auf einen Wert von möglichst unter 5 Prozent reduziert worden sein (vgl. Caliendo & Kopeining 2008). Darüber hinaus muss die nicht-parametrische Identifikation des Treatmenteffektes durch eine Überlappung der *Propensity Scores* (*Common Support*) der beiden Gruppen garantiert werden. Denn haben Treatment- und Kontrollgruppe eine ähnliche Verteilung der *Propensity Scores*, bedeutet dies, dass sie sich auch mit Blick auf die kontrollierten Kovariaten ähnlich sind (vgl. Rubin 2001).

Die Untersuchung der Mittelwertunterschiede (vgl. Tab. 1) offenbart, dass sich die SINUS-Schulen vor dem Matching hinsichtlich der Kontrollvariablen statistisch nachweisbar von den anderen Schulen im Datensatz nur mit Blick auf das Inspektionsjahr unterscheiden ( $p < 0.05$ ). Gleichwohl liegt der mittlere Standardisierte Bias vor dem Matching bei fast 30 und im Anschluss bei unter 3 Prozent; d. h. während die marginale Verteilung der Kovariaten vor dem Matching vergleichsweise hoch war, führt das Matching zu einer deutlichen Anpassung von Nicht-SINUS-Schulen an SINUS-Schulen mit Blick auf die kontrollierten Kovariaten.

Tabelle 1: Übersicht über die Kontrollvariablen vor und nach dem Matching und den standardisierten Bias

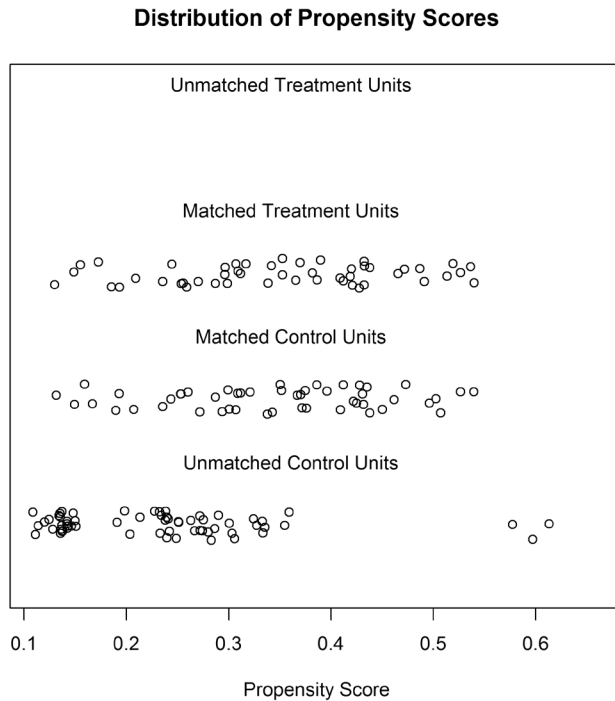
	vor Matching			nach Matching		
	SINUS (N=49)	ohne SINUS (N=112)	Signi- fikanz	SINUS (N=49)	ohne SINUS (N=49)	Signi- fikanz
Anzahl Schüler/-innen	131	121	p>0.10	131	132	p>0.10
Anzahl Lehrer/-innen	25	23	p>0.10	25	26	p>0.10
Inspektionsjahr 2007 (in %)	4,1	12,5		4,1	2,0	
Inspektionsjahr 2008 (in %)	6,1	14,3		6,1	8,2	
Inspektionsjahr 2009 (in %)	10,2	16,1		10,2	16,3	
Inspektionsjahr 2010 (in %)	18,4	16,1	p<0.05	18,4	18,4	p>0.10
Inspektionsjahr 2011 (in %)	38,8	27,7		38,8	40,8	
Inspektionsjahr 2012 (in %)	12,2	13,4		12,2	14,3	
Inspektionsjahr 2013 (in %)	10,2	0,0		10,2	0,0	
Sozialindex I (in %)	8,2	18,8		8,2	8,2	
Sozialindex II (in %)	16,3	17,0		16,3	20,4	
Sozialindex III (in %)	6,1	13,4		6,1	4,1	
Sozialindex IV (in %)	28,6	14,3	p>0.10	28,6	22,4	p>0.10
Sozialindex V (in %)	16,3	18,8		16,3	14,3	
Sozialindex VI (in %)	16,3	13,4		16,3	20,4	
Sozialindex nicht vorhanden (in %)	8,2	4,5		8,2	10,2	
Mittlerer Standardisierter Bias		29,23			1,54	

Auch weitere Gütemaße lassen darauf schließen, dass das Matching zu einer guten Anpassung der Kontroll- an die Treatmentgruppe geführt hat. So zeigt Abbildung 1 dass selbst in den Randbereichen der *Propensity-Score*-Verteilung Matching-Partner für die SINUS-Schulen gefunden werden konnten.

Und letztlich deuten sowohl der globale  $\chi^2$ -Test ( $p=0.93$ ) zur Prüfung der Stichprobenbalance nach Hansen und Bowers (2008) als auch die von Iacus, King und Porro (2009) zur Prüfung der Balanceoptimierung vorgeschlagene  $\mathcal{L}_1$ -Statistik ( $\Delta\mathcal{L}_1=0.06$ ) darauf hin, dass im vorliegenden Schuldatensatz für die gesamte Gruppe der SINUS-Schulen eine adäquat ausbalancierte Kontrollgruppe gefunden werden konnte und die Vergleichbarkeit von Treatment- und Kontrollgruppe durch das Matching hinsichtlich der genutzten Kovariaten merklich verbessert wurde.



Abbildung 1: Verteilung der Propensity Scores



## 5.2 Befunde zur Wirksamkeit des Programms SINUS an Grundschulen

### 5.2.1 Unterschiede in schulischen Prozessmerkmalen

Eine gute Schule zeichnet sich nach Ansicht der Hamburger Schulinspektion (vgl. IfBQ 2012) maßgeblich dadurch aus, dass unterrichtliche, erzieherische und organisatorische Prozesse gut gelingen. Welche Prozessmerkmale für eine gelingende Schul- und Unterrichtsentwicklung zentral sind, ist für Hamburger Schulen im Orientierungsrahmen Schulqualität (vgl. Behörde für Schule und Sport 2006) festgelegt. Der Orientierungsrahmen setzt Maßstäbe, an denen sich die Qualitätsentwicklung in Schulen orientieren soll. Auf diese Weise stellt er Kategorien für die Berichterstattung über Qualität bereit, die für die interne und externe Evaluation genutzt werden können. Der Orientierungsrahmen geht dabei davon aus, dass sich die Erfolge der Schule an ihren Wirkungen und Ergebnissen ablesen lassen. Ziel aller Qualitätsbemühungen einer Schule sind sichtbare Leistungen. Bedingungen dafür sind Verfahren der Führung und des Managements sowie Bildungs- und Er-



ziehungsprozesse, die für die Zielerreichung geeignet sind und deren Wirksamkeit empirisch nachgewiesen ist. Die Schulinspektion untersucht diese Verfahren und Prozesse. Die Schulinspektion Hamburg leitet ihre Bewertung der Qualität von Schule aus den 14 Qualitätsbereichen des Orientierungsrahmens ab, wobei Leistungen nicht bewertet werden. Die Bewertungen können dabei von „schwach“ über „eher schwach“ und „eher stark“ bis hin zu „stark“ reichen.

Tabelle 2: Bewertung der Qualitätsdimensionen an Schulen mit und ohne SINUS-Teilnahme (nach dem Matching)

Qualitätsdimension	SINUS	Nicht-SINUS	p
OR 1.1 Führung wahrnehmen	3,04	3,08	n.s.
OR 1.2 Personal entwickeln	2,60	2,40	n.s.
OR 1.3 Finanz- und Sachmittel gezielt einsetzen	3,27	3,12	n.s.
OR 1.4 Profil entwickeln und Rechenschaft ablegen	3,19	3,16	n.s.
OR 2.1 Das schuleigene Curriculum gestalten	2,08	1,98	n.s.
OR 2.2 Unterrichten, Lernen, Erziehen	2,96	2,76	<0.05
OR 2.3 Organisatorische Rahmenbedingungen sichern	3,35	3,20	n.s.
OR 2.4 Leistung beurteilen	2,42	2,38	n.s.
OR 2.5 Prozesse und Ergebnisse evaluieren	2,08	2,16	n.s.
OR 2.6 Förderkonzepte entwickeln	2,75	2,64	n.s.
OR 2.7 Beratungsangebote gestalten	2,83	2,82	n.s.
OR 2.8 Die Schulgemeinschaft beteiligen	2,88	2,84	n.s.
OR 3.1 Zufriedenes Personal, Schüler, Eltern, Betriebe	3,56	3,50	n.s.

OR = Orientierungsrahmen

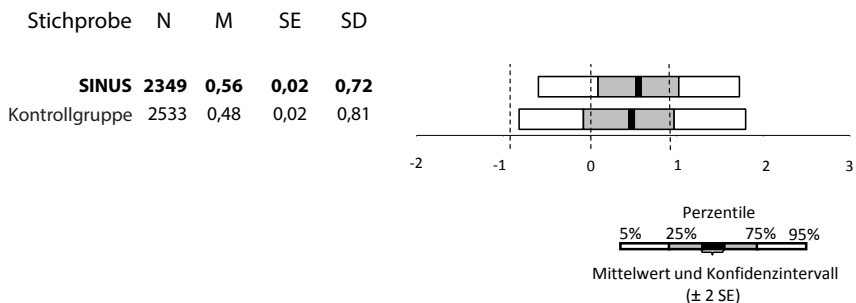
Diesbezüglich stellt sich die Frage, ob sich SINUS-Schulen von vergleichbaren Schulen nach Einschätzung der Schulinspektion systematisch in diesen Bereichen unterscheiden, wobei mit Blick auf die Qualität des Unterrichts insbesondere der Punkt 2.2 des Hamburger Orientierungsrahmens „Unterrichten, Lernen, Erziehen“ von Relevanz ist, da sich hierhinter die globale Einschätzung der Unterrichtsqualität an den inspizierten Schulen verbirgt, die durch die Inspektorinnen und Inspektoren der Schulinspektion auf Basis von Fragebogendaten, Interviews, Dokumentenanalysen und Unterrichtsbesuchen getroffen werden. Schaut man sich die Bewertungen der einzelnen Punkte an, so ist auffällig, dass sich Hamburger Grundschulen, die am Programm SINUS teilgenommen haben, ausschließlich in dem für das Programm relevanten Prozessbereich „Unterrichten, Lernen, Erziehen“ von vergleichbaren Schulen statistisch nachweisbar unterscheiden ( $p < 0.05$ , vgl. Tab. 2). In diesem Bereich liegt die Qualität an SINUS-Schulen nach Einschätzung der Inspektionsteams

um etwa 7 Prozent höher als an den Schulen der Kontrollgruppe. In allen anderen Bereichen, die die Schulinspektion einschätzt, unterscheiden sich Treatment- und Kontrollgruppe hingegen statistisch nicht nachweisbar voneinander ( $p > 0.10$ ). Auf Ebene der schulischen Prozesse ist demnach die Qualität des Unterrichts das einzige Unterscheidungsmerkmal zwischen SINUS und Nicht-SINUS-Schulen.

### 5.2.2 Unterschiede in Merkmalen der Unterrichtsqualität

Da sich die Einschätzungen der Inspektorinnen und Inspektoren zu diesem Punkt großteils aus den Unterrichtsbeobachtungen speisen, ist zu erwarten, dass auch hier differentielle Effekte nachweisbar sind. Eine Erwartung, die sich bestätigt (vgl. Abb. 2).

Abbildung 2: Qualität des Unterrichts an Schulen mit und ohne SINUS-Teilnahme (nach dem Matching)



So unterscheidet sich die Qualität des Unterrichts an SINUS-Schulen ( $M=0,56$ ) auch nach dem Matching noch statistisch signifikant ( $p < 0.01$ ) von derjenigen an vergleichbaren Schulen ( $M=0,48$ ), die nicht am Programm teilgenommen haben. Auffällig ist dabei vor allem auch, dass die Streuung der Qualität an den Programmschulen ( $SD=0,72$ ) selbst nach Kontrolle der Schulgröße deutlich geringer ausfällt als an vergleichbaren Grundschulen ( $SD=0,81$ ). An SINUS-Schulen ist demnach die Qualität des Unterrichts auf Schulebene homogener als an Schulen, die nicht an SINUS teilgenommen haben, jedoch unter vergleichbaren Bedingungen arbeiten. Betrachtet man Abbildung 2 genauer, so fällt darüber hinaus auf, dass sich die beiden Schulgruppen mit Blick auf die Verteilung vor allem im unteren Bereich unterscheiden. Extrem schwacher Unterricht ist an SINUS-Schulen demnach im Vergleich seltener zu finden als an vergleichbaren Schulen. Besonders deutlich wird dies, wenn man die Standards für Unterrichtsqualität der Schulinspektion anlegt.

Grundsätzlich verfehlen an allen Schulen des gematchten Datensatzes nur wenige Unterrichtssequenzen den Mindeststandard der Schulinspektion (N=141, 3%). Dabei reduziert die Teilnahme am SINUS-Programm das relative Risiko (RR), einen Unterricht an der Schule vorzufinden, bei dem es ausschließlich darum geht, das Lernklima und pädagogische Strukturen zu sichern – der also nicht lernwirksam ist –, um rund 56 Prozent bzw. erhöht die Chance (*Odds Ratio*, OR), zumindest den Mindeststandard an Unterrichtsqualität zu erreichen, um das 2,3-Fache. Mit Blick auf den Maximalstandard der Schulinspektion lassen sich hingegen keine bedeutsamen Unterschiede zwischen Treatment- und Kontrollgruppe feststellen.

Was dies praktisch bedeutet, veranschaulicht Tabelle 3. Dargestellt werden hier Merkmale der Unterrichtsqualität, die die Dimensionen effektiven Unterrichts, wie von Klieme (vgl. z.B. Klieme & Rackoczy 2008) beschrieben, charakterisieren. Bezogen auf dieses Modell, zeigt sich, dass die Klassenführung an allen in Hamburg inspizierten Grundschulen durchweg gut gelingt. Die Wahrscheinlichkeit, einen Unterricht zu finden, der im Rahmen der Schulinspektion diesbezüglich positiv eingeschätzt wird, beträgt an Schulen, die nicht am SINUS-Programm teilgenommen haben, 92 Prozent, an SINUS-Schulen sogar 94 Prozent ( $p<0.001$ ).

Tabelle 3:     Unterschiede in den Basisdimensionen effektiven Unterrichts an Schulen mit und ohne SINUS-Teilnahme (nach dem Matching)

	Klassenmanagement		Schülerorientierung		Kognitive Aktivierung		Lernautonomie	
Indikator-item	Die Arbeitsaufträge und Erklärungen sind angemessen, klar und präzise formuliert.		In den Unterricht sind der Erfahrungshorizont und/oder die Interessen der Schülerinnen und Schüler einbezogen.		Der Unterricht eröffnet Freiräume und ist nicht auf eine richtige Antwort fixiert.		Die Schülerinnen und Schüler werden angeregt/angeleitet, den Unterricht aktiv mitzugestalten, oder sie gestalten den Unterricht aktiv mit.	
	MW	p (MW>2.5)	MW	p (MW>2.5)	MW	p (MW>2.5)	MW	p (MW>2.5)
SINUS-Schulen	3,50	94%	3,00	73%	2,35	44%	1,93	28%
Kontrollgruppe	3,46	92%	2,97	70%	2,30	42%	1,90	27%

\*\*\* Mittelwertunterschied signifikant auf dem Niveau von 0.001; ns Mittelwertunterschied nicht signifikant

Auch hinsichtlich der Schülerorientierung resp. des unterstützenden Unterrichtsklimas zeigt sich, dass diese generell positiv wahrgenommen (Wahrscheinlichkeit einer positiven Bewertung 73 % vs. 70 %,  $p<0.001$ ) wird. Das Gleiche gilt für den Bereich der kognitiven Aktivierung. Die Wahrscheinlichkeit, dass dieses Merkmal sich an SINUS-Grundschulen positiv von vergleichbaren Schulen ab-

hebt, liegt auch hier nachweisbar höher ( $p < 0.001$ ), wobei sich jedoch zeigt, dass dieses Merkmal im Rahmen der Schulinspektion grundsätzlich mit unter 50-prozentiger Wahrscheinlichkeit positiv bewertet wird. Schaut man sich zu guter Letzt den Bereich der Lernautonomie an, so wird sichtbar, dass dieser im Rahmen der Unterrichtsbeobachtungen der Schulinspektion Hamburg selten positiv bewertet wird (Wahrscheinlichkeit einer positiven Bewertung 28 % vs. 27 %) und dass sich keinerlei Unterschiede zwischen SINUS- und vergleichbaren Grundschulen nachweisen lassen ( $p > 0.100$ ).

## 6. Diskussion und Ausblick

Es ist bekannt, dass wissenschaftsbasierte Schulentwicklungsmaßnahmen meist nachhaltige Effekte auf Schülerleistungen haben. Auch für das Programm SINUS an Grundschulen konnte kürzlich Vergleichbares mithilfe von Daten aus TIMSS 2011 festgestellt werden. Gleichwohl ist derzeit nur wenig über die Wirkungsmechanismen – also die vermittelnden Faktoren, auf Basis derer das Programm wirksam wird – bekannt. Befunde aus vorhergehenden Studien lassen erwarten, dass SINUS seine Wirkung primär über die Unterrichtsgestaltung entfaltet und sich der Unterricht an SINUS-Grundschulen somit von demjenigen an anderen Schulen, die nicht am SINUS-Programm teilnehmen, unterscheidet.

Mithilfe von Daten der Hamburger Schulinspektion, die im Zeitraum von 2006 bis 2013 Daten zur Unterrichtsqualität an allen Hamburger Schulen gesammelt hat und die den Anspruch hat, hiermit die Qualität des Unterrichts auf Schulebene empirisch verlässlich abbilden zu können, wurde diese Annahme geprüft. Da die Teilnahme am SINUS-Programm nicht zufällig erfolgte, wurde ein Quasi-Experiment durchgeführt, in dem a) verschiedene Prozessmerkmale effektiver Schulen, die im Rahmen der Inspektion durch Schulinspektorinnen und Schulinspektoren eingeschätzt werden, und b) der Unterricht an Hamburger Schulen, die an SINUS teilgenommen haben bzw. teilnehmen, mit denjenigen an Schulen verglichen wurden, die nicht an SINUS teilgenommen haben, jedoch unter vergleichbaren Bedingungen arbeiten. Um mögliche Selektionseffekte zu minimieren, wurde hierfür im Rahmen eines *Propensity Score Matching* die soziale Zusammensetzung der Schülerschaften, die Schulgrößen sowie das Inspektionsjahr der Schulen kontrolliert.

Die Befunde des Gruppenvergleiches bestätigen die bislang vorliegenden Befunde zu den Wirkungsmechanismen des Programms SINUS an Grundschulen. So konnte in einem ersten Schritt gezeigt werden, dass sich SINUS-Schulen von vergleichbaren Schulen mit Blick auf deren prozessuale Effektivitätsmerkmale – im Hamburger Orientierungsrahmen Qualitätsbereiche genannt – nach Einschätzung der Inspektionsteams ausschließlich im Bereich der Unterrichtsgestal-

tung statistisch nachweisbar voneinander unterscheiden. Entsprechend ließen sich weder Unterschiede in der Führung durch Schulleitungen, den organisatorischen Rahmenbedingungen des Unterrichts, der Nutzung von Evaluationskonzepten, der Beteiligung von Eltern und Schülern sowie der Zufriedenheit der Schulbeteiligten mit ihren Schulen nachweisen. Grundsätzlich scheint SINUS somit vor allem das intendierte Ziel einer verbesserten Unterrichtsentwicklung zu erreichen.

Was dies konkret bedeutet, wurde in einem zweiten Schritt gezeigt, in dem Daten aus Unterrichtsbeobachtungen der Schulinspektion genutzt wurden. Anhand von jeweils mehr als 2.000 Beobachtungen wird sichtbar, dass sich die Qualität des Unterrichts an SINUS-Schulen nachweisbar positiv von demjenigen an vergleichbaren Schulen abhebt. Im Mittel liegt die Qualität des Unterrichts an SINUS-Schulen rund 16 Prozent über derjenigen an vergleichbaren Schulen. Auffällig ist dabei weiterhin, dass, selbst bei Kontrolle der Schulgröße, die Variation der Unterrichtsqualität innerhalb der Stichproben unterschiedlich groß ist. So fällt die Streuung innerhalb der SINUS-Stichprobe rund 13 Prozent geringer aus als in der Kontrollgruppe. Und zu guter Letzt wird sichtbar, dass SINUS darüber hinaus dazu führt, dass Schwächen im Unterricht eliminiert werden. Unterricht, der den Mindeststandard an Unterrichtsqualität, wie ihn die Schulinspektion Hamburg definiert, nicht erreicht, ist an SINUS-Schulen sehr selten zu vorzufinden und findet dort weniger als halb so häufig statt wie an vergleichbaren Schulen.

Der Nachweis dieser drei Punkte ist dabei von besonderer Bedeutung, da die wenigen Studien, die Merkmale der Schul- mit denen der Unterrichtseffektivitätsforschung zusammenbringen (vgl. Ellett & Teddlie 2003), zeigen, dass der Unterricht an effektiven Schulen sich in genau diesen drei Bereichen von demjenigen an weniger effektiven Schulen unterscheidet. Unterricht an effektiven Schulen ist demnach gekennzeichnet durch a) eine generell höhere Unterrichtsqualität, b) eine geringere Streuung der Qualität innerhalb der Schule und c) dem Nichtvorhandensein von extrem schwachen Unterricht. Dies sind allesamt nachgewiesene Merkmale von SINUS-Schulen, was im Umkehrschluss darauf hindeutet, dass es sich bei SINUS-Schulen um Schulen mit einer vergleichsweise höheren Effektivität handelt, an denen Schülerinnen und Schüler höchstwahrscheinlich bessere Lernergebnisse erzielen können als an anderen Schulen.

Zusammenfassend lässt sich entsprechend feststellen, dass SINUS an Grundschulen grundsätzlich zu einer Effektivierung der Unterrichtspraxis führt und dass an diesen Schulen infolgedessen bessere Lernergebnisse aufseiten der Schülerinnen und Schüler zu erwarten sein sollten. Der Primäreffekt liegt dabei darin, dass Schwächen im Unterricht auf Ebene der Schule flächendeckend abgestellt werden. In der Konsequenz ermöglicht das Programm den Lehrkräften an den teilnehmenden Schulen, sich verstärkt auf die Motivations- und die Leistungsentwicklung der Schülerinnen und Schülern zu konzentrieren und den Unterricht zugewandt, fehlerfreundlich und unterstützend sowie kognitiv herausfordernd zu gestalten.

Besonders bedeutsam wird dieser Effekt, wenn man ihn auf das Gesamtsystem überträgt: So finden an Hamburgs Grundschulen in den Klassenstufen eins bis vier wöchentlich ca. 72.500<sup>1</sup> Unterrichtsstunden statt, wovon den Befunden der Inspektion zufolge etwa 2.200 (3 %) unterhalb des Mindeststandards liegen dürften. SINUS erhöht nun die Chance, dass der durch die Schulinspektion definierte Mindeststandard von Unterrichtsqualität erreicht wird, um das 2,3-fache ( $OR=2,3$ ,  $d=0.46$ ,  $r=0.23$ ) und hat entsprechend das Potenzial, den Anteil an Unterrichtsstunden pro Woche, in denen die Lernwirksamkeit nicht im Mittelpunkt steht, um rund ein Viertel zu senken. Das heißt, rund 500 Stunden pro Woche bzw. 20.000 Stunden pro Schuljahr, bei einem Schuljahr mit 40 Schulwochen, könnten durch eine flächendeckende Einführung des Programms an Grundschulen in Hamburg effektiver gestaltet und auf diesem Wege könnte die Wahrscheinlichkeit von Lernerfolgen aufseiten der Schülerinnen und Schüler erhöht werden.

Welche Folgen die Unterrichtspraxis an Hamburger SINUS-Schulen konkret für die Lernentwicklung von Schülerinnen und Schülern hat, kann die vorgelegte Studie leider nicht klären. Dies ist besonders bedauerlich, da Dalehefte et al. (2014) darauf verweisen, dass die von ihnen berichteten kausal interpretierten Befunde zur Wirksamkeit von SINUS auf die mathematische Kompetenz von Grundschülerinnen und Grundschulern nur bedingt haltbar seien, da kein experimentelles oder quasi-experimentelles Untersuchungsdesign genutzt wurde, das den Kompetenzzuwachs aufseiten der Schülerinnen und Schüler berücksichtigt.

Insofern hätte eine Verbindung von Prozess- und Ergebnisdaten im Rahmen der hier vorgelegten quasi-experimentellen Studie die Möglichkeit geboten, a) die Befunde aus TIMSS 2011 zu validieren und b) den Einfluss des veränderten Unterrichts an SINUS-Schulen auf die Kompetenzentwicklung von Schülerinnen und Schülern empirisch verlässlich abzuschätzen. Da flächendeckende Untersuchungen zur Lernentwicklung von Schülerinnen und Schülern in Hamburg erst zum Ende des ersten Inspektionszyklus eingeführt wurden, waren derartige Analysen hier jedoch nicht möglich. Zukünftig böte sich jedoch zur Klärung vergleichbarer Fragestellungen die Verbindung von Schulinspektionsdaten mit den ebenfalls in Hamburg vorhandenen Daten der KERMIT-Erhebungen (Kompetenzen ERMITteln) an, wobei auch in einem solchen Szenario etliche methodische Fallstricke zu berücksichtigen wären, um empirisch verlässliche Kausalaussagen zur Wirksamkeit von Schulentwicklungsmaßnahmen auf schulische Prozess- und Ergebnismerkmale zu gewährleisten.

1 Im Schuljahr 2013/2014 wurden an Hamburger Grundschulen 53.649 Schülerinnen und Schüler in den Klassenstufen eins bis vier in Klassen mit einer Größe von durchschnittlich 20 Personen unterrichtet. Die Anzahl der Grundstunden, die sie während der Grundschulzeit wöchentlich belegen müssen beträgt derzeit 108. Hieraus ergeben sich:  $\frac{53.649}{20} \times \frac{108}{4} = 72.426$  Unterrichtsstunden pro Schulwoche an Hamburgs Grundschulen.

## Literatur

- Behörde für Bildung und Sport (2006). *Orientierungsrahmen: Qualitätsentwicklung an Hamburger Schulen*. Hamburg: Behörde für Bildung und Sport.
- Borman, G., Hewes G., Overman, L. & Brown, S. (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. *Review of Educational Research*, 73 (2), 125–230.
- Caliendo, M. & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22 (1), 31–72.
- Dalehefte, I.-M., Wendt, H., Köller, O., Wagner, H., Pietsch, M., Döring, B., Fischer, C. & Bos, W. (2014). Bilanz von neun Jahren SINUS an deutschen Grundschulen: Evaluation im Rahmen der TIMSS 2011-Erhebung. *Zeitschrift für Pädagogik*, 60 (2), 245–263.
- Demuth, R., Walther, G. & Prenzel, M. (2011). *Unterricht entwickeln mit SINUS. 10 Module für den Mathematik- und Sachunterricht in der Grundschule*. Seelze: Klett-Kallmeyer.
- Ellet, C.D. & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17 (1), 101–128.
- Fischer, C. (2005). Rolle von Schulleitungen bei innovativen Projekten. Aus Sicht der Grundschule. *SchulVerwaltung. Zeitschrift für Schulleitung, Schulaufsicht und Schulkultur Ausgabe Niedersachsen und Schleswig-Holstein*, 9, 251–253.
- Fischer, C., Kobarg, M., Dalehefte I.-M. & Trepke, F. (2013). Ein Unterrichtsentwicklungsprogramm wissenschaftlich begleiten. Anlage und Hintergründe des Forschungsdesigns. *Psychologie in Erziehung und Unterricht*, 60 (1), 26–31.
- Fischer, C. & Prenzel, M. (2007). Wie Lehrkräfte bei der Qualitätssicherung unterstützt werden können. Das Beispiel SINUS-Transfer Grundschule. *Journal für Schulentwicklung*, 11 (2), 23–39.
- Fischer, C., Rieck, K. & Lobemeier, K.R. (2008). Mit Logbüchern dokumentieren und reflektieren. Das Beispiel SINUS-Transfer Grundschule. In E. M. Lankes (Hrsg.). *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 73–86). Münster: Waxmann.
- Fischer, C., Rieck, K. & Dedekind, B. (2009): SINUS-Transfer Grundschule. Lehrkräfte verändern ihren Mathematikunterricht und ihren naturwissenschaftlichen Sachunterricht an Grundschulen – (wie) geht das? *MNU PRIMAR*, 2, 44–49.
- Hansen, B.B. & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23, 219–236.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? *Pädagogik*, 58 (2), 42–45.
- Hill, C. J., Bloom, H. S., Black, A. R. & Lipsey, M. W., (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2 (3), 172–177.
- Iacus, S.M., King, G. & Porro, G. (2009). CEM: Coarsened exact matching software. *Journal of Statistical Software*, 30, 1–27.
- IfBM (2011). *Jahresbericht der Schulinspektion Hamburg. Schuljahr 2009–2010*. Hamburg: Institut für Bildungsmonitoring.



- IfBQ (2012). *Jahresbericht der Schulinspektion Hamburg. Schuljahr 2010–2011*. Hamburg: Institut für Bildungsmonitoring und Qualitätsentwicklung.
- Klieme, E. (2011). *Standards für die Unterrichtsqualität. Kann es das geben?* Vortrag an der Universität Zürich, 20. November 2011.
- Klieme, E. & Rackoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222–237.
- Legewie, J. (2012). Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64 (1), 123 – 153.
- Leithwood, K. & Jantzi, D. (2009). A Review of Empirical Evidence about School Size Effects: A Policy Perspective. *Review of Educational Research*, 79 (1), 464–490.
- Meyer, H. (2008). Disput über aktuelle Probleme und Aufgaben der Didaktik. In M. A. Meyer, M. Prenzel & S. Hellekamps (Hrsg.), *Perspektiven der Didaktik (Zeitschrift für Erziehungswissenschaft: Sonderheft 9)*, S. 77–86). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meyer, H. (2011). *Die Rolle der Schulleitung bei der Unterrichtsentwicklung*. Kiel: IPN.
- Meyer, H. & Klapper, A. (2006). Unterrichtsstandards für ein kompetenzorientiertes Lernen und Lehren. In R. Hinz & B. Schumacher (Hrsg.), *Auf den Anfang kommt es an: Kompetenzen entwickeln – Kompetenzen stärken* (S. 89–108). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Morgan, S. L. & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Ostermeier, C. (2005). *Kooperative Qualitätsentwicklung in Schulnetzwerken: Eine empirische Studie am Beispiel des BLK-Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts (SINUS)“*. Münster: Waxmann.
- Pietsch, M. (2010). Evaluation von Unterrichtsstandards. *Zeitschrift für Erziehungswissenschaft*, 13 (1), 121–148.
- Pietsch, M. (2011). Fehlende Daten bei Unterrichtsbeobachtungen: Eine Sensitivitätsanalyse anhand von Daten der Schulinspektion Hamburg. *Empirische Pädagogik*, 25 (1), 47–87.
- Pietsch, M., Jahnke, N. & Mohr, I. (2014). Führt Schulinspektion zu besseren Schülerleistungen? Difference-in-Difference-Studien zu Effekten der Schulinspektion Hamburg auf Lernzuwächse und Leistungstrends. *Zeitschrift für Pädagogik*, 60 (3), 446–470.
- Pietsch, M., Schulze, P., Schnack, J. & Krause, M. (2011). Elaborierte Rückmeldungen zur Qualität von Unterricht. Über empirisch abgesicherte Bezugsnormen als Grundlage für die Weiterentwicklung von Unterricht und Schule. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektionen in Deutschland – Eine Zwischenbilanz aus empirischer Sicht* (S. 193–216). Münster: Waxmann.
- Prenzel, M., Friedrich, A. & Stadler, M. (2009). *Von SINUS lernen. Wie Unterrichtsentwicklung gelingt*. Seelze: Klett-Kallmeyer.



- Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C. & Seidel, T. (2005). Wie schneiden SINUS-Schulen bei PISA ab? Ergebnisse der Evaluation eines Modellversuchsprogramms. *Zeitschrift für Erziehungswissenschaft*, 8 (4), 540–561.
- Rosenbaum, P. & Rubin, D. (1983). The Central Role of Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70 (1), 41–55.
- Rosenbaum, P. & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33–38.
- Seidel T. & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77 (4), 454–499.
- Rubin, D. (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.
- Stadler, M. & Meentzen, U. (2010). Wie Lehrkräfte bei der Reflexion über ihren Unterricht unterstützt werden können. Das Fachgruppenportfolio im Programm SINUS-Transfer. In F. H. Müller (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung* (S. 161–173). Münster: Waxmann.
- Stralla, M. (2009). *Die Unterrichtsbeobachtungen im Rahmen der deutschen Schulinspektion. Analyse des Kerninstruments zur Beurteilung der Schulqualität* (Unveröffentlichte Diplomarbeit). Berlin: Freie Universität Berlin.
- Thoemmes, F. (2014). *Propensity score matching in SPSS*. Verfügbar unter: <http://sourceforge.net/projects/psmspss/> [Zugriff am 30.03.2015]
- Thoemmes, F. & Kim, E. S. (2011). A Systematic Review of Propensity Score Methods in the Social Sciences. *Multivariate Behavioral Research*, 46, 90–118